IOP Publishing Journal of Physics: Condensed Matter
J. Phys.: Condens. Matter **30** (2018) 144002 (8pp) https://doi.org/10.1088/1361-648X/aab0c6

Self consistent field theory of virus assembly

Siyu Li[1,4], **Henri Orland**[2,3] **and Roya Zandi**[1]

[1] Department of Physics and Astronomy, University of California, Riverside, CA 92521, United States of America
[2] Institut de Physique Théorique, CEA-Saclay, CEA, F-91191 Gif-sur-Yvette, France
[3] Beijing Computational Science Research Center, No.10 East Xibeiwang Road, Haidan District, Beijing 100193, People's Republic of China

E-mail: sli032@ucr.edu

Received 17 December 2017, revised 12 February 2018
Accepted for publication 20 February 2018
Published 12 March 2018

Abstract

The ground state dominance approximation (GSDA) has been extensively used to study the assembly of viral shells. In this work we employ the self-consistent field theory (SCFT) to investigate the adsorption of RNA onto positively charged spherical viral shells and examine the conditions when GSDA does not apply and SCFT has to be used to obtain a reliable solution. We find that there are two regimes in which GSDA does work. First, when the genomic RNA length is long enough compared to the capsid radius, and second, when the interaction between the genome and capsid is so strong that the genome is basically localized next to the wall. We find that for the case in which RNA is more or less distributed uniformly in the shell, regardless of the length of RNA, GSDA is not a good approximation. We observe that as the polymer–shell interaction becomes stronger, the energy gap between the ground state and first excited state increases and thus GSDA becomes a better approximation. We also present our results corresponding to the genome persistence length obtained through the tangent–tangent correlation length and show that it is zero in case of GSDA but is equal to the inverse of the energy gap when using SCFT.

Keywords: persistence length, field theory, virus assembly

Supplementary material for this article is available online

(Some figures may appear in colour only in the online journal)

1. Introduction

Viruses have evolved to optimize the feat of genome packaging inside a nano-shell called the capsid, built from several copies of either one or a few different types of proteins. Quite remarkably, under many circumstances the capsid proteins of single-stranded RNA viruses can assemble spontaneously [1–9] around the cognate and non-cognate RNAs and other negatively charged cargos [7, 10–13]. It is widely accepted that the electrostatic interaction is the main driving force for the assembly [2–6, 14–16] and it is this feature that has made viruses ideal for various bio-nanotechnological applications including gene therapy and drug delivery.

Despite their great interest in biological and industrial applications, the physical factors contributing to the efficient assembly and stability of virus particles are not well understood [17, 18]. The difficulty emerges from the considerable number of variables in the system including the genome charge density, the persistence length, the surface geometry and the charge density of surface charges. The adsorption of genome to the inner wall of capsid, the interplay between long-range electrostatic and short-range excluded volume interactions and the issue of chain connectivity make the understanding of the problem quite challenging. The presence of salt makes the adsorption process even more complicated. The salt ions can screen the electrostatic interaction between the charges and modify the persistence length of the genome leading to a change in the profile of the genome in the capsid.

[4] Author to whom any correspondence should be addressed.

Because of the difficulties noted above, in all previous studies on the encapsidation of viral genome by capsid proteins, the ground state dominance approximation, in which only the lowest energy eigenstate of a negatively charged genome inside a positively charged viral shell is considered, has been exclusively used [19–27]. In this paper, we investigate the validity of the ground state dominance approximation (GSDA) in different regimes as a function of salt concentration, genome charge density and surface charge density. Note that viral RNA is relatively long compared to the capsid inner radius. For example for many plant viruses, RNA is about 3000 nucleotides while the inner capsid radius is around 10 nm [28]. While it is well-known that GSDA works well for long chains [29], in many recent virus assembly experiments short pieces of RNA have been systematically employed, to study the impact of genome length on the virus stability and formation [30]. Thus the time is ripe to explore the conditions under which GSDA does not apply and self consistent field theory has to be solved to obtain the correct solution. Comparing the solutions of the self-consistent field theory (SCFT) and GSDA shows that GSDA is less accurate when the interaction of genome with the capsid wall is weak even if the genome is long.

The paper is organized as follows. In the next section, we introduce the model and all the relevant equations. In section 3, we present our results, and we discuss the impact on the genome profile of the capsid charge density, salt concentration and polymer length and charge density in section 4. Finally, also in section 4, we present our conclusion and summarize our findings.

2. Theory

In order to calculate the free energy of a virus particle in a salt solution, we model the capsid as a positively charged shell, in which a negatively charged flexible linear polymer (genomic RNA) is confined. Defining by N the number of monomers, N_+ the number of salt cations and N_- the number of salt anions, the partition function of the system can be written as

$$Z = \sum_i^{N_+}\sum_i^{N_-} \frac{1}{N_+!}\frac{1}{N_-!} e^{\beta\mu N_+} e^{\beta\mu N_-} \int \mathcal{D}r_i^+ \mathcal{D}r_i^- \mathcal{D}r_s \exp\left\{-\frac{3}{2a^2}\int_0^N \mathrm{d}s\dot{r}_s^2 - \frac{1}{2}\int \mathrm{d}r\mathrm{d}r'\hat{\rho}_m(r)u(r-r')\hat{\rho}_m(r') - \frac{\beta}{2}\int \mathrm{d}r\mathrm{d}r'\hat{\rho}_c(r)v_c(r-r')\hat{\rho}_c(r')\right\} \quad (1)$$

where a is the Kuhn length of the monomers. We assume that the salt is monovalent (charge e per ion), and the charge per monomer is τ. The monomer density $\hat{\rho}_m(r)$ and the charge density $\hat{\rho}_c(r)$ are given by

$$\hat{\rho}_m(r) = \int_0^N \mathrm{d}s\delta(r - r_s) \quad (2)$$

$$\hat{\rho}_c(r) = \rho_0(r) + \tau\int_0^N \delta(r - r_s)\mathrm{d}s + e\left(\sum_i^{N^+}\delta(r - r_i^+) - \sum_i^{N^-}\delta(r - r_i^-)\right) \quad (3)$$

where $\rho_0(r)$ denotes the charge density of the viral shell. In equation (1), the term $u(r) = u_0\delta(r)$ represents Edwards's excluded volume interaction, and $v_c(r) = 1/4\pi\epsilon r$ is the Coulomb interaction between the charges, where ϵ is the dielectric permitivity of the solvent.

2.1. Self consistent field theory

To obtain the genome profile inside the virus capsid, we use self-consistent field theory (SCFT [31]) and the grand canonical ensemble for the salt ions with their fugacity λ corresponding to the concentration of salt ions in the bulk. Performing two Hubbard–Stratonovich transformations and introducing the excluded volume field $w(r)$ and the electrostatic interaction field $\phi(r)$ (see the online supplementary material (stacks.iop.org/JPhysCM/30/144002/mmedia)), equation (1) simplifies to

$$\mathcal{Z} = \int \mathcal{D}w(r)\mathcal{D}\phi(r)\, e^{\log Q - \int \mathrm{d}r\{\frac{1}{2u_0}w^2(r) + \frac{\beta\epsilon}{2}(\nabla\phi(r))^2 - 2\lambda\cosh(\mathrm{i}\beta e\phi(r)) + \mathrm{i}\beta\rho_0(r)\phi(r)\}}$$

where Q denotes the partition function for a single chain

$$Q = \int \mathcal{D}r_s e^{-\frac{3}{2a^2}\int_0^N \mathrm{d}s\dot{r}_s^2 - \mathrm{i}\int \mathrm{d}r\hat{\rho}_m(r)[w(r) + \beta\tau\phi(r)]}. \quad (4)$$

The self-consistent field theory equations are obtained by performing the saddle-point approximation on the two integration fields w and ϕ, see the supplementary material. The equations are

$$w(r) = u_0\rho_m(r) \quad (5)$$

$$-\epsilon\nabla^2\phi = -2\lambda e\sinh(\beta e\phi(r)) + \rho_0(r) + \beta\tau\rho_m(r) \quad (6)$$

where

$$\rho_m(r) = \int_0^N \mathrm{d}s\, q(r, N - s)q(r, s) \quad (7)$$

is the monomer concentration at point r. Equation (6) is the Poisson–Boltzmann equation for the charged monomers–salt ions system [32].

In equation (7), we have introduced the propagator $q(r, s)$, which is proportional to the probability for a chain of length s to start at any point in the viral shell and to end at point r [33]. It satisfies the SCFT (diffusion) equation [34],

$$\frac{\partial q(r, s)}{\partial s} = \frac{a^2}{6}\nabla^2 q(r, s) - V(r)q(r, s) \quad (8)$$

$$V(r) = w(r) + \beta\tau\phi(r) \quad (9)$$

with the following boundary condition

$$q(r,0) = \frac{1}{\sqrt{Q}} \tag{10}$$

for r anywhere in the virus shell. The single chain partition function Q is given in equation (4) and is determined through the normalization condition on $q(r,s)$

$$\int_0^N \mathrm{d}r q(r,N-s)q(r,s) = 1 \text{ for any } s. \tag{11}$$

Note that the SCFT equation (8) can also be written as an imaginary time Schrödinger equation in the form

$$\frac{\partial q(r,s)}{\partial s} = -Hq(r,s) \tag{12}$$

with the Hamiltonian H given by

$$H = -\frac{a^2}{6}\nabla^2 + V(r). \tag{13}$$

Once we obtain the propagator q then we can calculate the chain persistence length or stiffness as explained in the next section.

2.2. Persistence length

Polymers may have some bending rigidity or stiffness, due either to their intrinsic mechanical structure or to the Coulombic interaction between charged monomers, which has a tendency to rigidify the chain. This stiffness results in a strong correlation between the orientation of successive monomers. Eventually, at large separations, the directions of monomers become uncorrelated. The persistence length of a polymer is the correlation length of the tangents to the chain [34, 35]. It is the typical distance over which the orientation of monomers becomes uncorrelated. The chain can be viewed as a set of independent fragments of length equal to their persistence length.

In order to compute the persistence length, we calculate the correlation function of tangents to the chain

$$C(s,s') = \langle \dot{r}(s)\dot{r}(s')\rangle. \tag{14}$$

We show in the supplementary material that within the SCFT, this correlation function can be expressed as

$$C(s,s') = \frac{a^4}{9}\int \mathrm{d}r\mathrm{d}r' \left(\frac{\partial}{\partial r}q(r,N-s)\right)\left(\frac{\partial}{\partial r'}q(r',s')\right) \times \langle r|\mathrm{e}^{-(s-s')H}|r'\rangle \tag{15}$$

where we assumed that $s > s'$. In this equation, for brevity we have used the standard quantum mechanical representation for the matrix elements of the evolution operator, see for example equations (S5), (S9) and (S28) in the supplementary material.

For large separation $s - s' \gg 1$, this function behaves as

$$C(s,s') \approx \mathrm{e}^{-(s-s')/l_p} \tag{16}$$

where by the above definition, l_p is the persistence length of the chain.

2.3. Ground state dominance approximation

The set of non-linear partial differential equations given in equations (6) and (8) are very tedious to solve. In the case of a confined chain, or more generally for a system with a gap in the energy spectrum of the Hamiltonian H, it is convenient to use the so-called ground state dominance approximation as noted in the introduction. This approximation consists of expanding the propagator q (equation (8)) in terms of the eigenfunctions of the Hamiltonian H. We thus write

$$q(r,s) = \sum_{k=0}^{\infty} \mathrm{e}^{-E_k s} q_k \psi_k(r) \tag{17}$$

where $\{E_k, \psi_k(r), k = 0,1,2,\ldots\}$ are the set of normalized eigenvalues and eigenstates of H, respectively,

$$\begin{aligned} H\psi_k(r) &= E_k\psi_k(r) \\ \int \mathrm{d}r\, \psi_k^2(r) &= 1. \end{aligned} \tag{18}$$

Using the boundary condition equation (10), we find

$$q_k = \frac{1}{\sqrt{Q}}\int \mathrm{d}r\psi_k(r) \tag{19}$$

with

$$Q = \sum_{k=0}^{\infty} \mathrm{e}^{-NE_k}\left(\int \mathrm{d}r\psi_k(r)\right)^2. \tag{20}$$

We assume that the eigenvalues are ordered as $E_0 < E_1 < \ldots < E_k < \ldots$. When the energy gap between the ground state E_0 and the first excited state E_1 is large, the ground state dominates the expansion equation (17) and we may write

$$q(r,s) = \mathrm{e}^{-E_0 s}\left(q_0\psi_0(r) + \mathrm{e}^{-s\Delta}R(r,s)\right) \tag{21}$$

where $\Delta = E_1 - E_0$ is the energy gap, and the function $R(r,s)$ is the remainder of the expansion. When $s\Delta \gg 1$, the second term above becomes exponentially negligible, and we may write

$$q(r,s) = \mathrm{e}^{-E_0 s} q_0 \psi_0(r) \tag{22}$$

and then equations (20), (19) and (7) become respectively equal to

$$Q = \mathrm{e}^{-NE_0}\left(\int \mathrm{d}r\psi_0(r)\right)^2 \tag{23}$$

$$q_0 = \mathrm{e}^{NE_0/2} \tag{24}$$

$$\rho_m(r) = N\psi_0^2(r). \tag{25}$$

The Poisson–Boltzmann (equation (6)) and diffusion (equation (8)) equations then become

$$\begin{aligned} &-\epsilon\nabla^2\phi = -2\lambda e\sinh(\beta\phi) + N\tau\psi_0(r)^2 + \rho_0 \\ &-\frac{a^2}{6}\nabla^2\psi_0(r) + Nu_0\psi_0(r)^3 + \beta\tau\phi(r)\psi_0(r) = E_0\psi_0(r) \end{aligned} \tag{26}$$

and the energy E_0 is determined so that ψ_0 is normalized as

$$\int \mathrm{d}r\, \psi_0^2(r) = 1. \tag{27}$$

Similarly, we can compute the correlation function equation (15) within the GSDA. Using equation (24) and the fact that

$$\langle r|\mathrm{e}^{-(s-s')H}|r'\rangle = \mathrm{e}^{-(s-s')E_0}\psi_0(r)\psi_0(r') \tag{28}$$

in ground state dominance (GSD), we obtain

$$\begin{aligned} C(s,s') &= \frac{a^4}{9}\left(\int \mathrm{d}r\psi_0(r)\frac{\partial\psi_0}{\partial r}\right)^2 \\ &\equiv 0 \end{aligned} \tag{29}$$

since the integral is identically 0. We conclude that in the GSDA, the persistence length vanishes. In order to have a non-vanishing persistence length, we need to include more than the ground state in the eigenstate expansion of all quantities. Including the next leading order term (first excited state with energy E_1 and wave function ψ_1), we obtain (see the supplementary material)

$$C(s,s') \approx A_1\mathrm{e}^{-|s-s'|\Delta} + A_2\mathrm{e}^{-(N-|s-s'|)\Delta} \tag{30}$$

which shows that the persistence length is the inverse of the gap

$$l_p = \frac{1}{\Delta}. \tag{31}$$

The persistence length can be computed using the GSDA as it follows: having solved the GSD equation (26), we know E_0, $\psi_0(r)$ and $\phi(r)$ from which we can calculate $q(r,s)$ and the Hamiltonian H. We can then compute the first excited state of H with energy E_1, and then the persistence length l_p from equation (31).

3. Results

Due to the complexity of the problem, we numerically solve the non-linear coupled equations given in equations (6) and (8). We consider two different cases for the interaction of genome with the capsid. First we study the adsorption of the chain to the capsid inner wall in the absence of the electrostatic interactions, as explained in section 3.1 below. This way we decrease the number of parameters in the system, which helps us to gain some insights before solving the full problem. Then in section 3.2, we assume that both the capsid and chain are charged in salt solution.

3.1. Confined RNA with adsorption on capsid

We consider the confined RNA adsorbed on the capsid wall with no electrostatic interaction present. Thus, the external field (V_{ext}) in equation (8) contains only the excluded volume interaction between monomers(u_0), with an extra attraction from the capsid γ_s. To solve the diffusion equation (8) with this surface term is not trivial, the strategy we introduce therefore is the effective boundary condition [36]:

$$\left[\frac{\partial}{\partial r}q(r,s) - \kappa q(r,s)\right]_{r=R} = 0 \tag{32}$$

where κ^{-1} is the extrapolation length and is proportional to the inverse of γ_s.

We employ both SCFT and GSDA to solve the problem of a chain confined in an adsorbing spherical shell. To obtain the exact solutions for SCFT, we solve equations (6) and (8) recursively until conditions in equations (10), (11) and (32) are satisfied. We employ the Crank–Nicolson scheme and Broyden method [37, 38] to solve the relevant equations. For the approximative solutions of GSD, we operate on the coupled nonlinear equations (equation (26)) with finite element method and deal with the convergence issue using the Newton method.

The results of our calculations are presented in figure 1, which shows the confined RNA density profile as a function of r, the distance from the shell center, for various extrapolation length (κ^{-1}). The goal is to compare our findings obtained through the GSDA and SCFT methods for both short and long RNA. The dashed lines in figure 1 are obtained using GSDA while solid lines are calculated based on the SCFT method. As illustrated in the figure, GSD only makes a good approximation for long chains and/or short extrapolation lengths (strong adsorption regime or large κ). With short RNA or long extrapolation length (weak adsorption regime), GSDA profile deviates considerably from self-consistent profile. As illustrated in figure 1, for $N = 5000$ regardless of the strength of interaction κ^{-1}, the solutions of GSDA and SCFT match almost perfectly and completely cover each other. However, the agreement between the two methods becomes less for $N = 100$ and small values of κ. In the next section, we investigate the impact of electrostatic interaction on the profile of RNA inside the capsid.

3.2. Confined RNA with electrostatic interaction

Since RNA acts like a negatively charged polyelectrolyte in solution, we need to take into consideration the electrostatic interactions term $\beta\tau\phi(r)$ given in equation (8). We assume that positive charges on the capsid are uniformly distributed. The coulombic interaction does usually overwhelm other forces responsible for the adsorption of chain to the wall, so instead of applying the Robin boundary condition (equation (32)) as in section 3.1, we use the Dirichlet boundary condition ($q(R,s) = 0$) for monomer density by assuming the V_{ext} is infinity beyond the capsid wall. The physical basis for this assumption is that the RNA monomer has stiffness, and the excluded volume interaction between the capsid wall and the RNA is such that the density of RNA could never sit at the wall.

We then solve equations (6) and (8) to obtain the RNA density through both GSDA and SCFT methods. The genome concentration profiles are shown in figure 2 for various RNA length (total monomer number), capsid charge density, chain charge density and salt concentrations. As expected, there is

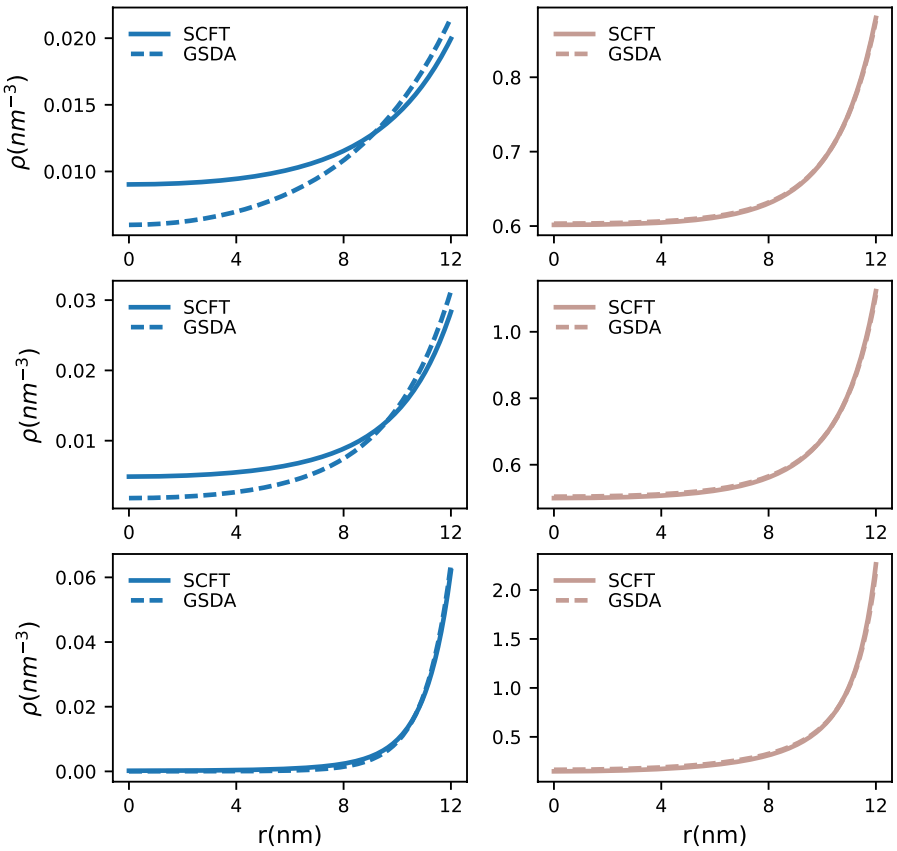

Figure 1. Confined RNA density profile versus r the distance from the capsid center for various extrapolation lengths, $\kappa^{-1} = 10.0, 5.0, 2.0$ nm for top to the bottom of the figure. The total monomer number is $N = 100$ (left), $N = 5000$ (right).

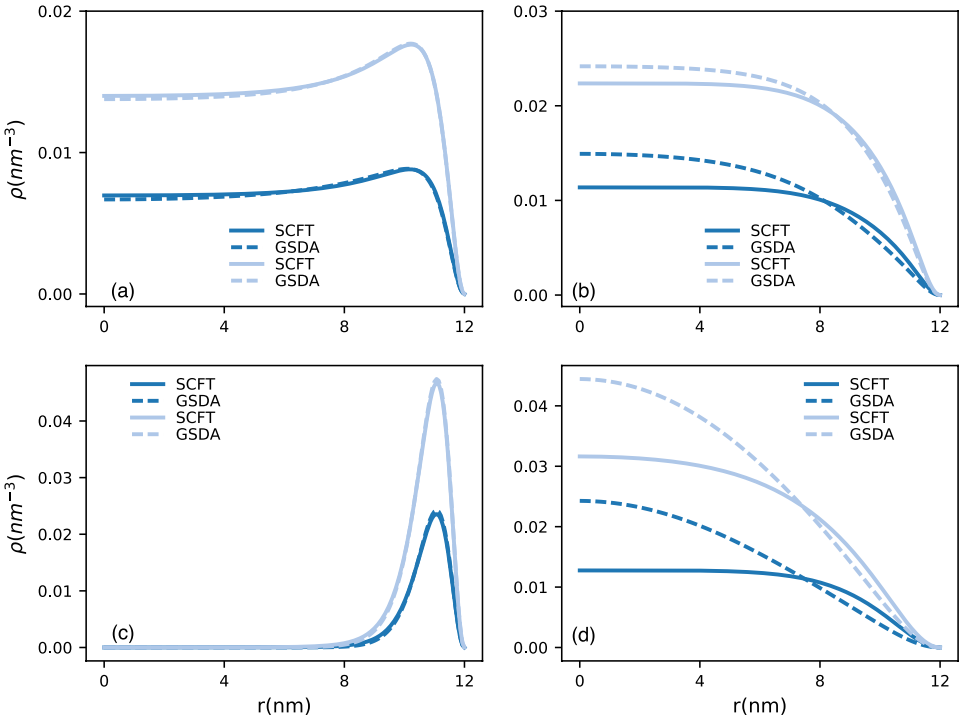

Figure 2. Confined RNA concentration profiles with various RNA lengths $N = 50$ (darker), $N = 100$ (lighter) under SCFT calculation (solid lines) and GSDA (dashed lines) with (a) linear chain charge density $\tau = -1.0e$, capsid surface charge density $\sigma = 0.8e$ nm^{-2} and salt concentration $\lambda = 500$ mM; (b) $\tau = -1.0e$, $\sigma = 0.4e$ nm^{-2} $\lambda = 500$ mM; (c) $\tau = -1.0e$, $\sigma = 0.4e$ nm^{-2}, $\lambda = 100$ mM. (d) $\tau = -0.1e$, $\sigma = 0.4e$ nm^{-2}, $\lambda = 100$ mM; Other parameters used are Kuhn length $a = 1$ nm, excluded volume $u_0 = 0.05$ nm^3, and capsid radius $R = 12$ nm.

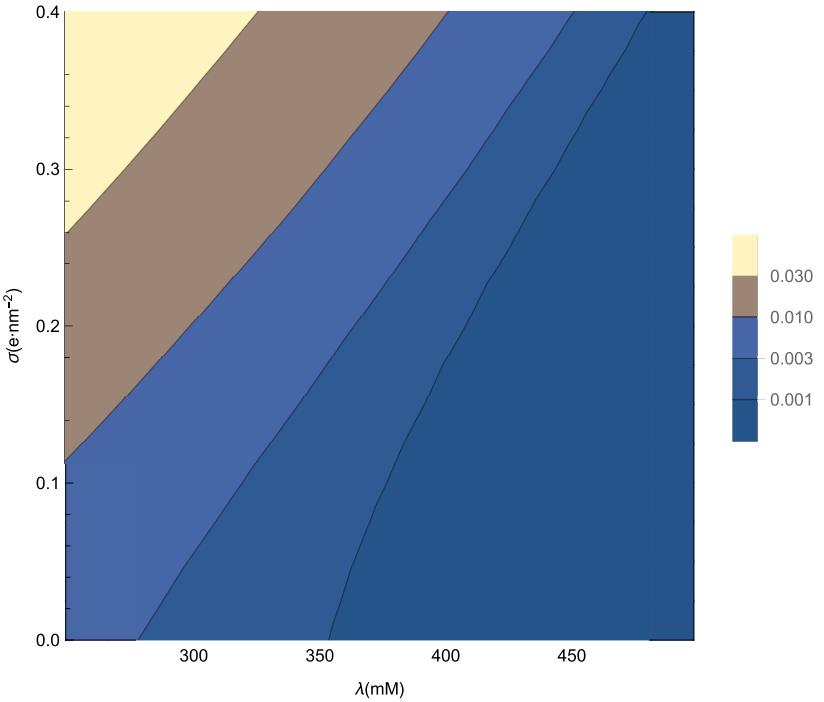

Figure 3. Genome excess phase diagram with respect to salt concentration and capsid surface charge density. The white shade corresponds to the region with the maximum genome density and black to the depletion regime next to the wall. Other parameters used are $N = 500$, $a = 1\,\text{nm}$, $u_0 = 0.05\,\text{nm}^3$, $R = 12\,\text{nm}$.

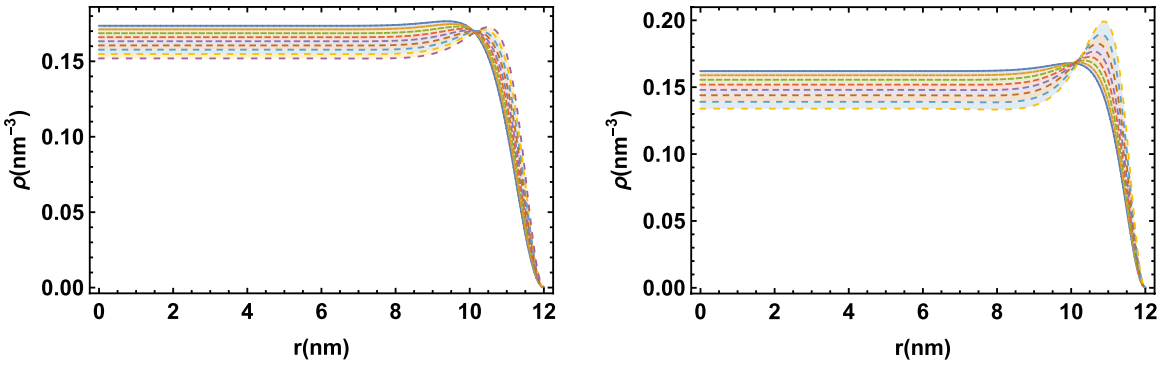

Figure 4. Genome density profile for $N = 1000$ and (a) various surface charge densities (0–0.4e nm^{-2}) with salt concentration $\lambda = 400$ mM; (b) various salt concentrations (250–500 mM) with fixed surface charge $\sigma = 0.4e$ nm^{-2}. Other parameters correspond to Kuhn length $a = 1\,\text{nm}$, excluded volume $u_0 = 0.05\,\text{nm}^3$, capsid radius $R = 12\,\text{nm}$.

alway a perfect match between GSDA and SCFT for longer RNA (large N), while for short RNA (small N), the energy gap becomes considerable and important, with ground state less dominant in the whole expansion series (equation (17)) and the GSDA becomes less valid.

We also find that the stronger the electrostatic interaction due to the higher capsid surface charge density or genome linear charge density, the better the GSDA and SCFT results agree with each other. Figure 2(a) shows that regardless of length of genome, at high surface charge density, GSDA and SCFT give the same results. Note, as we decrease the surface charge density, their difference becomes noticeable, as illustrated in figure 2(b). However, with lower salt concentration for the same surface charge density as in figure 2(b), the difference between the two methods once again becomes negligible, figure 2(c). Quite interestingly as we decrease the chain linear charge density even at low salt, we find again that the agreement between the two models becomes detectable, figure 2(d).

All results presented above show that GSDA is less valid when genome localizes close to the center. To this end, we investigate this transition point where the wall attraction becomes so weak that depletion shows up, corresponding to the disappearance of the genome peaks in graphs of figures 2(a), (b) and also figures 4(a), (b). We calculate the excess genome at the wall by integrating the genome peak area, which is proportional to adsorbed monomers. Then we investigate the impact of the salt concentration and surface charge density on the adsorption-depletion transition. The resulting phase diagram is illustrated in figure 3. The white shade in the figure corresponds to the maximum adsorption. As the color gets darker, less genome is adsorbed to the wall.

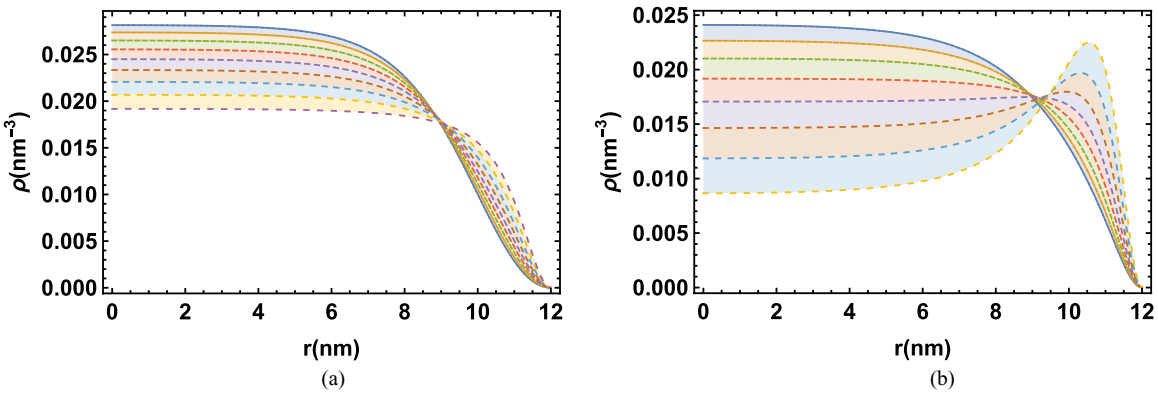

Figure 5. Genome density profile for $N = 100$ and (a) various surface charge densities (0–0.4e nm^{-2}) with salt concentration $\lambda = 400$ mM; (b) various salt concentrations (250–500 mM) with fixed surface charge $\sigma = 0.4e$ nm^{-2}. Other parameters correspond to $a = 1$ nm, $u_0 = 0.05$ nm^3, $R = 12$ nm.

In the darkest region there is no adsorption. The line separating the darkest region indicates the onset of the depletion transition.

Figure 4 describes the genome profile details for two different cases. For a fixed salt concentration but varying surface charge density ($\sigma = 0{-}0.4e$ nm^{-2}) we observe that the peak next to the wall slowly disappears as the capsid charge density decreases and most of the genome becomes localized at the center, figure 4(a). Similar behavior is displayed in figure 4(b) for a fixed surface charge density but various salt concentrations. Figures 4(a) and (b) together tell us that the higher salt concentration, or the lower surface density charge, causes genome to stay away from the capsid wall and to localize toward the center, constructing the region where GSDA is not valid any more.

4. Discussion and summary

The results of previous sections show that the GSDA validity depends on the genome localization: when the genome is absorbed on the wall, GSDA works perfectly, however when the adsorption becomes weaker and the genome starts moving to the center, GSDA stops being reliable. Figure 2 illustrates this statement, where perfect match between GSDA and SCFT is obtained in lower salt concentration and higher surface charge (localized genome); significant deviation appears at higher salt concentration and lower surface charge in which case the genome is delocalized. The same effect is observed for the linear charge density of short genomes.

For longer genomes with 500 monomers or more, the difference is almost undetectable. Quite interestingly, the effect of the electrostatic interaction range and strength, salt concentration and surface charge density in section 3.2 is similar to that of the extrapolation length in section 3.1. While low salt concentration (longer Debye length, strong attraction) and high surface charge correspond to larger κ, high salt concentration (short Debye length, weak attraction) and low surface charge on the contrary correspond to small κ in which case the GSDA does not work well as illustrated in figure 1.

Another important difference arising from using the GSDA and SCFT approaches corresponds to the tangent–tangent correlation function or persistence length of the polymer. While the persistence length obtained through GSDA is zero, the persistence length calculated using SCFT is inversely proportional to the energy gap between the ground state and the first excited state, equation (31). The vanishing persistence length in GSDA is due to the fact that the chain constraint or connectivity is absent, and all monomers are independent. In the case of SCFT, the persistence length increases with the length of genome until it saturates to a finite value. Then indeed, as N increases, $l_p \ll N$, explaining again why GSDA becomes more and more valid as the length of the genome increases.

While the persistence length corresponds to the stiffness of the polymer, there is another important length scale in the problem but it is associated with the adsorption of polymer on the inner shell of the capsid. The adsorption of polymers to flat surfaces have been thoroughly studied, but the adsorption to spherical shells is less understood [39–42]. In case of flat surfaces, the Edward's correlation length determines the distance from the wall over which the adsorption layer decays. It goes as $\xi \sim 1/\sqrt{u_0\phi_B}$, with u_0 the strength of the excluded volume and ϕ_B the bulk polymer density.

The situation studied in this paper is more complex due to the confinement of the polymer inside a spherical capsid in the presence of electrostatics. Quite interestingly, figures 4(a) and (b) show there is a point around $r = 10$ nm where all the curves cross. According to the figure, the location of the crossing point does not depend on the salt concentration and capsid charge density. Since the capsid is a closed shell, we cannot define the bulk density in this problem. However, ϕ_B is related to the number of monomers in the capsid. Figures 5(a) and (b) illustrate the genome profiles for the same parameters as in figures 4(a) and (b) respectively but using a shorter genome length. The genome length is $N = 100$ and $N = 1000$ in figures 4 and 5, respectively. As illustrated in figure 5 all the plots again meet at a particular point but the position of the crossing point is moved compared to figure 4. It is interesting that despite different capsid charge density and salt concentration, all curves again meet at a unique single distance from the wall.

We also checked the position of the crossing point as a function of the excluded volume interaction expressed through the

Edward's correlation length $\xi \sim 1/\sqrt{u_0\phi_B}$. Our numerical results did not show any dependence of the crossing point on the strength of the excluded volume interaction. This is probably due to the fact that ϕ_B in this problem is not really the bulk density and depends on the excluded volume interaction and might cancel the impact of the excluded volume interaction. Although we cannot provide a closed form formula for the Edward's correlation length, it is interesting that all points meet at one single point and this point is independent of the capsid charge density, salt concentration and the polymer excluded volume interaction but depends on the length of genome.

In summary, in this paper we investigated the validity of GSDA for studying the profile of genomes in viral shells because of the extensive usage of GSDA in the literature in describing the process of virus assembly and stability. We found that for small RNA segments employed in recent experiments or for *in vitro* assembly studies with mutated capsid proteins carrying lower charge density [15, 30, 43], the GSDA deviates from the accurate results obtained through SCFT methods. Otherwise, native RNA viruses are long enough compared to the radius of the capsid and as such GSDA is good enough to explain different experimental observations and there is no need to solve tedious self-consistent equations. Our results showed that the narrower the region RNA occupies and the stronger the genome–capsid interaction is, the larger the energy gap, and hence the better GSDA works.

Acknowledgments

The authors would like to thank Xingkun Man for useful discussions. This work was supported by the National Science Foundation through Grant No. DMR -1719550 (RZ).

ORCID iDs

Siyu Li https://orcid.org/0000-0002-4326-7560

References

[1] Perlmutter J D, Qiao C and Hagan M F 2013 *eLife* **2** e00632
[2] Sikkema F D, Comellas-Aragones M, Fokkink R G, Verduin B J M, Cornelissen J and Nolte R J M 2007 *Org. Biomol. Chem.* **5** 54
[3] Ren Y P, Wong S M and Lim L Y 2006 *J. Gen. Virol.* **87** 2749
[4] Ni P, Wang Z, Ma X, Das N C, Sokol P, Chiu W, Dragnea B, Hagan M and Kao C C 2012 *J. Mol. Biol.* **419** 284
[5] Losdorfer Bozic A, Siber A and Podgornik R 2013 *J. Biol. Phys.* **39** 215
[6] Zlotnick A, Aldrich R, Johnson J M, Ceres P and Young M J 2000 *Virology* **277** 450
[7] Sun J *et al* 2007 *Proc. Natl Acad. Sci. USA* **104** 1354
[8] Ning J, Erdemci-Tandogan G, Yufenyuy E L, Wagner J, Himes B A, Zhao G, Aiken C, Zandi R and Zhang P 2016 *Nat. Commun.* **7** 13689
[9] Luque A, Zandi R and Reguera D 2010 *Proc. Natl Acad. Sci.* **107** 5323
[10] Kusters R, Lin H-K, Zandi R, Tsvetkova I, Dragnea B and van der Schoot P 2015 *J. Phys. Chem.* B **119** 1869
[11] Hagan M F and Zandi R 2016 *Curr. Opin. Virol.* **18** 36
[12] Bruinsma R F, Comas-Garcia M, Garmann R F and Grosberg A Y 2016 *Phys. Rev.* E **93** 032405
[13] Stockley P G *et al* 2013 *J. Biol. Phys.* **39** 277
[14] Lin H-K, van der Schoot P and Zandi R 2012 *Phys. Biol.* **9** 066004
[15] Sivanandam V, Mathews D, Garmann R, Erdemci-Tandogan G, Zandi R and Rao A L N 2016 *Sci. Rep.* **6** 26328
[16] Tao H, Rui Z and Shklovskii B I 2008 *Physica* A **387** 3059
[17] Wagner J and Zandi R 2015 *Biophys. J.* **109** 956
[18] Wagner J, Erdemci-Tandogan G and Zandi R 2015 *J. Phys.: Condens. Matter* **27** 495101
[19] van der Schoot P and Bruinsma R 2005 *Phys. Rev.* E **71** 061928
[20] Zandi R and van der Schoot P 2009 *Biophys. J.* **96** 9
[21] Erdemci-Tandogan G, Wagner J, van der Schoot P, Podgornik R and Zandi R 2014 *Phys. Rev.* E **89** 032707
[22] Erdemci-Tandogan G, Wagner J, van der Schoot P, Podgornik R and Zandi R 2016 *Phys. Rev.* E **94** 022408
[23] Li S, Erdemci-Tandogan G, Wagner J, Van Der Schoot P and Zandi R 2017 *Phys. Rev.* E **96** 022401
[24] Li S, Erdemci-Tandogan G, van der Schoot P and Zandi R 2018 *J. Phys.: Condens. Matter* **30** 044002
[25] Siber A and Podgornik R 2008 *Phys. Rev.* E **78** 051915
[26] Siber A, Zandi R and Podgornik R 2010 *Phys. Rev.* E **81** 051919
[27] Erdemci-Tandogan G, Orland H and Zandi R 2017 *Phys. Rev. Lett.* **119** 188102
[28] Comas-Garcia M, Cadena-Nava R D, Rao A L N, Knobler C M and Gelbart W M 2012 *J. Virol.* **86** 12271
[29] de Gennes P-G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)
[30] Rayaprolu V, Moore A, Wang J C-Y, Goh B C, Perilla J R, Zlotnick A and Mukhopadhyay S 2017 *J. Phys.: Condens. Matter* **29** 484003
[31] Edwards S F 1965 *Proc. Phys. Soc.* **85** 613
[32] Borukhov I, Andelman D and Orland H 1998 *Eur. Phys. J.* B **5** 869
[33] Fredrickson G 2006 *The Equilibrium Theory of Inhomogeneous Polymers* vol 134 (Oxford: Oxford University Press)
[34] Doi M and Edwards S F 1986 *The Theory of Polymer Dynamics* (*International Series of Monographs on Physics* vol 73) (Oxford: Oxford Science Publications)
[35] Abels J A, Moreno-Herrero F, van der Heijden T, Dekker C and Dekker N H 2005 *Biophys. J.* **88** 2737
[36] Ji H and Hone D 1988 *Macromolecules* **21** 2600
[37] Man X, Yang S, Yan D and Shi A C 2008 *Macromolecules* **41** 5451
[38] Wang Q, Taniguchi T and Fredrickson G H 2004 *J. Phys. Chem.* B **108** 6733
[39] Joanny J F 1999 *Eur. Phys. J.* B **9** 117
[40] Pincus P A, Sandroff C J and Witten T A 1984 *J. Phys.* **45** 725
[41] Hone D, Ji H and Pincus P A 1987 *Macromolecules* **20** 2543
[42] Eisenriegler E, Kremer K and Binder K 1982 *J. Chem. Phys.* **77** 6296
[43] Brasch M, Voets I K, Koay M S T and Cornelissen J J L M 2013 *Faraday Discuss.* **166** 47